AD-A071 108 CENTER FOR NAVAL ANALYSES ALEXANDRIA VA MARINE CORP--ETC F/G 12/1
CALCULATING SAMPLE SIZE REQUIREMENTS FOR MILITARY UNIT INSPECTI--ETC(U)
MAR 79 L J GRIKE N00014-76-C-0001

UNCLASSIFIED CRC-379 NL

| OF |
AD
A071108

END
DATE
FILMED
8-79
DDC

CRC 379 / March 1979

CALCULATING SAMPLE SIZE REQUIREMENTS FOR MILITARY UNIT INSPECTIONS

By Leo J. Grike

Approved for public release; distribution unlimited

CNA *Marine Corps Operations Analysis Group*

CENTER FOR NAVAL ANALYSES

2000 North Beauregard Street, Alexandria, Virginia 22311

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14

1. REPORT NUMBER
CRC-379

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

6

4. TITLE (and Subtitle)
Calculating Sample Size Requirements for Military Unit Inspections

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

10

7. AUTHOR(s)
Leo J. Grike

15

8. CONTRACT OR GRANT NUMBER(s)
N00014-76-C-0001

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Center for Naval Analyses
2000 N. Beauregard Street
Alexandria, Virginia 22311

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Deputy Chief of Staff (RD&S)
Headquarters, Marine Corps
Washington, D.C. 20380

11

12. REPORT DATE
March 1979

13. NUMBER OF PAGES
48

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

12 54p

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

9 Research contribution,

18. SUPPLEMENTARY NOTES

This Research Contribution does not necessarily represent the opinion of the Department of the Navy.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

computer programs, hypergeometric functions, inspection, military forces, normal approximation, rates, sampling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Computer programs for calculating sample size requirements are presented and illustrated. Given the desired significance level and the desired probability that the sample will achieve that level, sample size requirements are calculated for determining that a unit rating:

- exceeds a criterion on level (with replacement);
- is less than a criterion level (with replacement);

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

401 327

20

- is between two criterion levels (with replacement); and
- exceeds a criterion level (without replacement).

The last calculation is done exactly (with the hypergeometric distribution) and also with a normal approximation. In all cases, the required sample size is calculated as a function of the true unit rating.

CENTER FOR NAVAL ANALYSES
2000 North Beauregard Street, Alexandria, Virginia 22311 Telephone: (703) 998-3500

CNA

8 June 1979

MEMORANDUM FOR DISTRIBUTION

Subj: Center for Naval Analyses Research Contribution 379

Encl: (1) CRC 379, "Calculating Sample Size Requirements for Military Unit Inspections," by Leo J. Grike, March 1979

1. Enclosure (1) is forwarded as a matter of possible interest.

2. This Research Contribution gives methods for estimating, a priori, sample size requirements for comparisons to percentage criteria. Though developed for military inspections, the methods are useful for any situation in which the sample size must be chosen in advance.

3. Research Contributions are distributed for their potential value in other studies and analyses. They do not necessarily represent the opinion of the U.S. Marine Corps or the Department of the Navy.

ROBERT J. CORN
Director
Marine Corps Operations
Analysis Group

Distribution List:
Reverse page

Subj: Center for Naval Analyses Research Contribution 379

DISTRIBUTION LIST

Department of the Navy

SNDL Part I:

21A1	CINCLANTFLT
21A2	CINCPACFLT
21A3	CINCUSNAVEUR
24J1	CG FMFLANT
24J2	CG FMFPAC

SNDL Part II:

A1	ASST SECNAV (MRA&L)
A5	NAVMILPERSCOM
A6	DC/S, Manpower
A6	DC/S, RD&S
A6	Inspector Gen'l
B3	NAT'L DEF UNIV
B3	AFSC
FB44	HUMRESMANCEN, San Diego
FB44	HUMRESMANCEN, Pearl
FF38	USNA, Nimitz Library
FF44	NAVWARCOL
FF48	HUMRESMANCEN
FKA6A16	NAVPERSRANDCEN
FL3	DODCI
FT1	CNET
FT73	NAVPGSCOL
FT87	HUMRESMANSCOL
V12	CG MCDEC
V12	MCDEC, Breckenridge Lib

OpNav: Op-09BH, Op-01

Other

Ass't Secretary of Defense, Program Analysis & Evaluation (2 copies)
Defense Documentation Center (12 copies)
USAF Air University
Institute for Defense Analyses
Human Resource Research Organization
The Rand Corporation

TABLE OF CONTENTS

I. SUMMARY

BACKGROUND

The Inspector-General of the Marine Corps requested assistance in putting the sampling procedures of his inspectors on a sounder statistical basis. His principal requirement to achieve this is to start with random or nearly-random samples. This aspect of the problem is being treated as a thesis topic at the Naval Postgraduate School. This paper examines the question of sample size requirements. Recommended sample sizes are derived for sampling with and without replacement. Two types of rating systems (both of which use percentages) are treated: one with two ratings, and one with more than two ratings. Actual systems have more than two ratings, but practical sample sizes may require treating these as two-level systems.

METHOD OF ANALYSIS

We must first establish what is possible, in principle, to determine in advance of an inspection. We assume that the assignment of individual ratings (physical fitness test results, rifle qualification scores, etc.) are without error. This means that a repeated test would give an unaltered result, which is a reasonable assumption since the criteria used are broad: pass or fail, qualify or fail to qualify. The situation is then analogous to taking samples of black and white balls from a box containing a known number of balls, but with unknown proportions of black and white. In general, no sample size less than 100 percent will guarantee a sample that will justify a given statistical confidence in the results. What _is_ possible is to calculate a sample size such that, given a true proportion p and unit size N, the significance level of the sample will be better than some level α a specified percentage of the time .

For discussion, we take 90 percent as a practical minimal success probability, and 0.1 as a practical significance level. Given that we have correctly estimated the true population proportion, our calculated sample sizes will then give ratings significant at the 0.1 level or better (compared to some standard) in 90 percent of the cases. If the true population proportion is closer to the standard than we assumed, we will be successful less than 90 percent of the time; if farther, the frequency will be over 90 percent. Since the ratings are continuous percentages, the actual unit performance could be extremely close to the standard, so that the sample size tends to infinity for sampling with replacement and to unit size without replacement. The results of our calculations will show the tradeoff between sample size and the range of unit performances likely to be properly and confidently classified.

In practice, the inspection planner can make a low estimate of unit performance, and look up the sample size required to place that unit properly "most" of the time. If practicable, a sample of that size is taken; if not, the largest practicable sample is used.

II. DETERMINING SAMPLE SIZE REQUIREMENTS FOR SPECIFIC LEVELS OF INSPECTION

SAMPLE SIZES WITH TWO-LEVEL RATINGS (WITH REPLACEMENT)

Computationally, the simplest case is for sampling with replacement in a two-level rating system. Actual rating systems will tend to have more than two levels, but it will be shown that practical sample sizes do not allow finer distinctions than pass or fail. Therefore, the two-level case tends to be the interesting one. We assume that there is some criterion level, below which unit performance is unsatisfactory. In our examples, this percentage is 84.5.

The method[1] used is based on the Student t-distribution, with the percentage data normalized by an arcsin transformation. The sample size S required is given by:

$$S \geq \frac{820.8[t_{\alpha} + t_{2(1-P)}]^2}{(\text{Arcsin}\sqrt{p} - \text{Arcsin}\sqrt{p_o})^2} ,$$

where α is the significance level chosen, P is the probability of attaining a level α, p is the assumed true unit performance (a fraction), and p_o is the criterion performance level. The constant is obtained from the variance of the arcsin function in degrees squared. An "infinite" number of degrees of freedom is assumed. The numerator is not a function of p or p_o, and can easily be tabulated for a useful range of P and α. (Reference 1 tabulates twice the numerator, which is appropriate when both percentages are measured. The last two columns of the table in reference 1, labeled .01 and .001, respectively, should be labeled .02 and .01.)

Figure 1 and table 1 show the results of this calculation for success probabilities ranging from 80 to 99 percent. Appendix A gives a computer program for calculating these sample sizes.

Note that the formula applies to the difference between two percentages, either of which could be the greater of the two. Except for terminology, the same calculations serve to determine the sample size required to show that true unit performance is below some criterion level. A program for this calculation is also given in appendix A.

---

[1] See reference 1, pages 609-10.

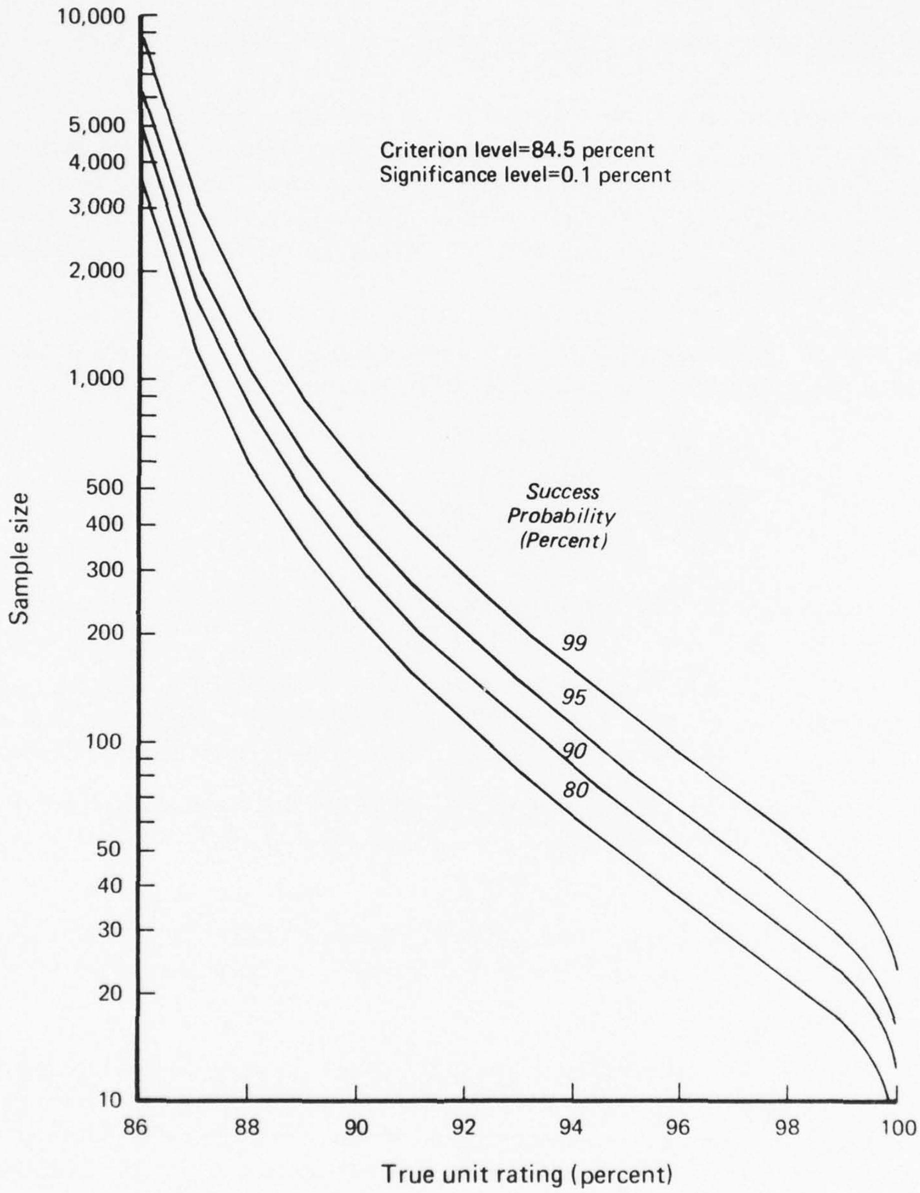

FIG. 1: SAMPLE SIZE REQUIRED WITH REPLACEMENT TO SHOW UNIT TO BE ABOVE CRITERION LEVEL

TABLE 1

SAMPLE SIZES REQUIRED WITH REPLACEMENT TO SHOW THAT A UNIT RATING IS ABOVE A CRITERION LEVEL[a]

True unit rating	Success probability (percent)				
	80	85	90	95	99
0.85	31967	37292	44278	55959	81538
0.86	3454	4029	4784	6046	8810
0.87	1207	1408	1672	2113	3078
0.88	596	696	826	1044	1521
0.89	349	407	483	610	889
0.9	225	262	312	394	574
0.91	155	180	214	271	395
0.92	111	130	154	195	284
0.93	82	96	114	144	210
0.94	63	73	87	110	160
0.95	48	56	67	84	123
0.96	37	44	52	65	95
0.97	29	34	40	51	74
0.98	22	26	31	39	57
0.99	17	19	23	29	43
1	9	11	13	17	24

[a]Criterion level = 84.5 percent
Significance level = 0.1

SAMPLE SIZES WITH THREE-LEVEL RATINGS (WITH REPLACEMENT)

With three or more levels, it may be desired to determine that a unit's true performance is within a given rating band, meaning that it is simultaneously above a lower criterion level and below an upper level. Consider the distribution of sample ratings for some true unit performance. For a given sample size, some set of the possible sample ratings will justify the statement that at some level of significance the true rating is above the criterion level. This set of ratings will constitute a fraction F_{above} of the possible sample outcomes, and will cluster as far as possible from the lower criterion rating. A similar statement can be made with respect to the upper criterion level, where the fraction will be called F_{below}. The area of the sample distribution for which the statement can be made about <u>both</u> criterion levels simultaneously is equal to the area of either F, less the area of that F not occupied by the other F (figure 2). We call this area F_{inside}.

$$F_{inside} = F_{above} - (1-F_{below}),$$

$$= F_{below} - (1-F_{above}).$$

or

$$F_{inside} = F_{below} + F_{above} - 1.$$

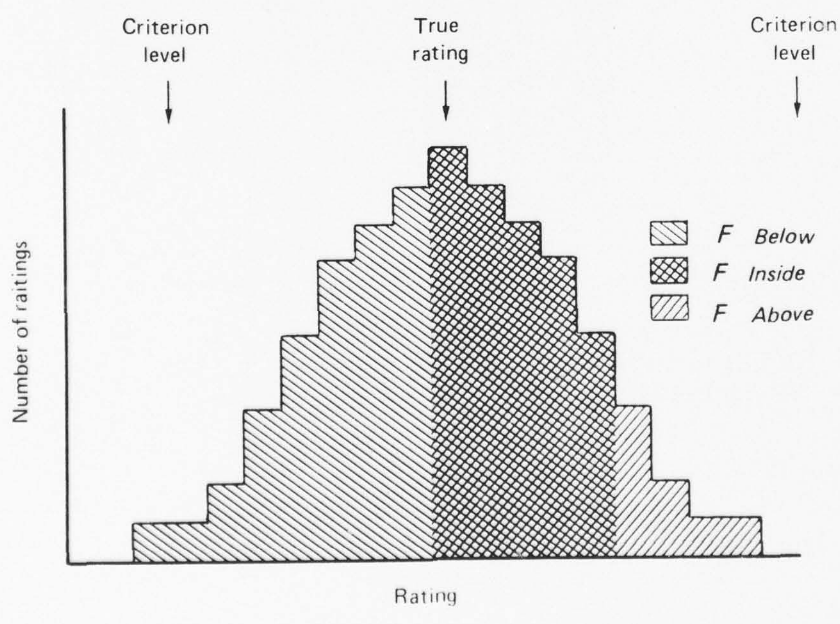

FIG. 2: SAMPLE RATINGS DISTRIBUTION

For example, to obtain a 90-percent chance of a successful sample, we need $F_{inside} = 0.9$. The sum of the other Fs is then 1.9.

Available tables permit calculation of probabilities as high as 99.5 percent. These probabilities and the following procedure were used to determine the sample sizes for which F_{inside} has some specified value, say 0.9. First, for some assumed unit rating, calculate the sample sizes required for a range of success probabilities, with respect to one of the two criterion levels. Using these sample sizes, calculate the t-values associated with a comparison of the unit rating with the other criterion level; that is, calculate $t_{2(1-P)}$. From a fifth-degree polynomial fit of P versus $t_{2(1-P)}$, calculate the success probabilities with respect to the second criterion level. (All Ps greater than .995 are taken equal to one.) The sum of the two success probabilities, less one, is the success probability for a sample being inside the rating band. The range of F_{above}, say, gives a range of F_{below} and, therefore, a range of F_{inside}. The last are logarithmically interpolated to obtain the sample sizes required for various selected success probabilities. Some of these sample sizes are shown in figure 3 and table 2. Appendix A gives a program for calculating these probabilities and sample sizes.

SAMPLING WITH REPLACEMENT

With replacement, the required sample size can easily exceed the unit size. In these cases, we could simply sample the entire unit. However, with our assumptions about the repeatability of test scores, it is not necessarily true that the entire unit must frequently be measured. The distribution of the number of different individuals who would be sampled in a given case (with given unit and sample sizes) can be calculated. If the mean of this distribution is appreciably less than the unit size, an appropriate sample can be obtained by counting randomly selected individuals more than once. As an example of the typical sample size savings that can be obtained in this way, we will calculate the <u>most likely</u> number of different individuals in a sample. Feller (reference 2, page 92) shows how to calculate exactly the distribution of the number of different individuals. The formula, however, requires double precision calculations on the Burroughs B6700 computer to handle sample and unit sizes much over 30. Even double precision calculations will not permit the range of unit sizes required. Fortunately, a drastic simplification is available. First, the large-number limit of the exact formula is the Poisson distrubution (reference 2, pages 93-94). Second, for our purpose of finding the peak of the distribution, the exact and the Poisson distrubutions coincide above unit and sample sizes of 15 (determined empirically). Finally, the parameter of the Poisson distribution gives (almost exactly) the most likely number of individuals <u>not</u> in the sample, so that no Poisson terms need be calculated. The parameter is

$$\lambda = N \, EXP[-R/N],$$

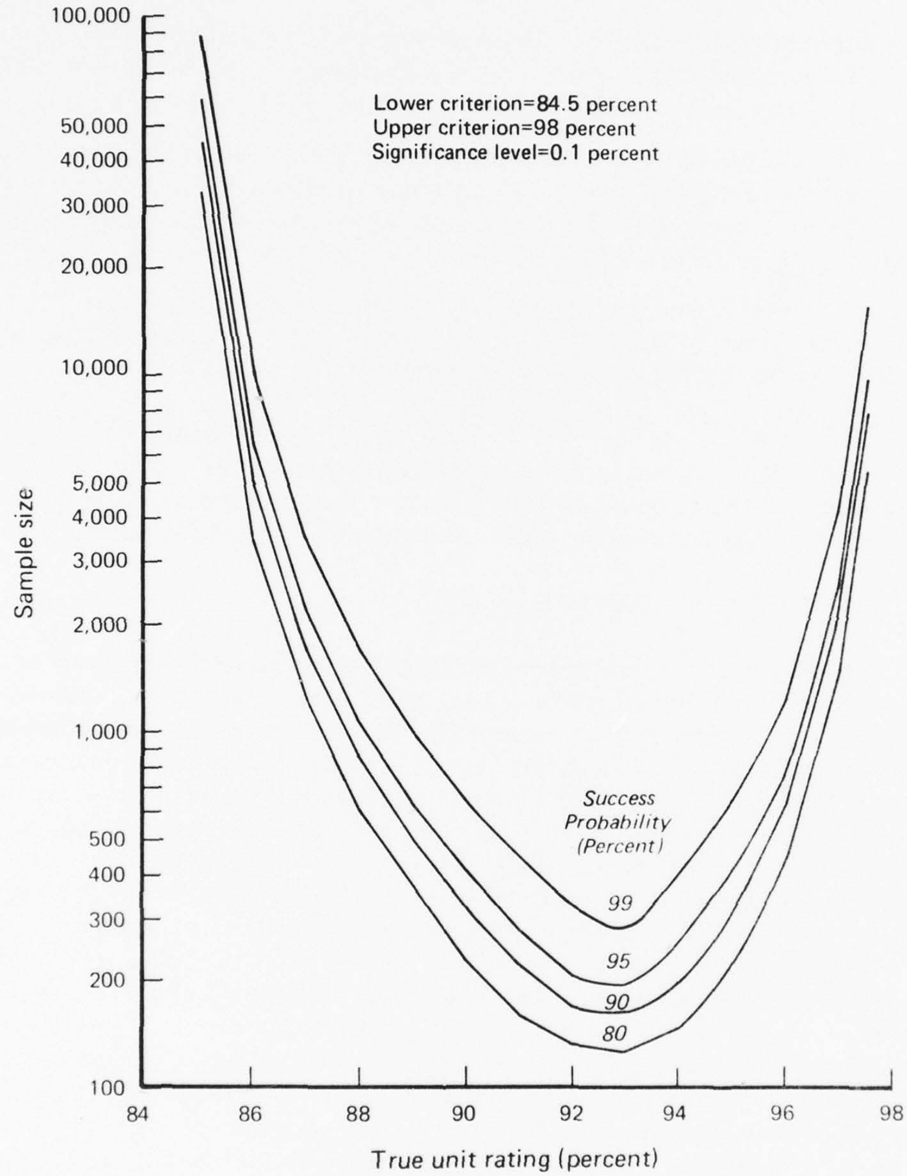

FIG. 3: SAMPLE SIZE REQUIRED WITH REPLACEMENT TO SHOW UNIT TO BE BETWEEN CRITERION LEVELS

TABLE 2

SAMPLE SIZES REQUIRED WITH REPLACEMENT
TO SHOW THAT A UNIT RATING IS IN A RATING LEVEL[a]

True unit Rating	Success probability (percent)				
	80	85	90	95	99
85	32520	38425	45358	58045	92099
86	3514	4151	4900	6272	9951
87	1228	1451	1712	2191	3477
88	606	717	846	1083	1718
89	355	419	495	633	1005
90	229	271	319	409	648
91	159	187	219	281	446
92	132	147	169	206	320
93	127	140	161	193	281
94	147	170	197	252	400
95	225	266	313	402	637
96	445	526	621	794	1260
97	1518	1794	2118	2710	4299

[a]Lower criterion level = 84.5 percent
Upper criterion level = 98 percent
Significance level = 0.1

where N is the unit size, and R is the sample size (without replacement). Forward differences suggest that λ is exactly the most likely number of missing individuals; backward differences suggest that λ is one unit too large, and empirical calculations show that λ is either correct or one unit too small. In practice, since the distributions tend to be very broad, we can take the required sample size (number of different individuals) to be $N-\lambda$.

Some exact answers are given in table 3. The bottom line shows the requirements for units of infinite size. The other entries show the numbers of different individuals in a typical sample of the size shown at the bottom of the table, taken without replacement from a unit of the size given at the left. These entries are limited in size by both the unit size and sample size for an infinite unit. The sample size savings implied by these entries are shown in table 4. Table 4 indicates that, while there are combinations of unit size and true ratings for which weighting the sample can produce significant savings in sample size, the savings are usually too small to justify the complexities.

SAMPLE SIZES WITH TWO-LEVEL RATINGS (WITHOUT REPLACEMENT)

The traditional way to take inspection samples is without replacement. Conceptually, this method is also the simplest: No one is sampled or counted more than once. In this case, the sampling distribution is given by the hypergeometric distribution (reference 3, page 193); when the population is much larger than the sample, the binomial distribution is adequate.

The hypergeometric distribution has four parameters and requires computation of many factorials. The terms of the distribution are, therefore, not very convenient for calculation or tabulation. Nevertheless, a computer makes it feasible to use the hypergeometric distribution for this problem. The method is somewhat complicated.

There are two classes of hypergeometric distributions involved. We first assume a unit (population) size, a desired significance level, and a desired success probability. Suppose the latter two to be 0.1 and 90 percent, respectively. We also require a criterion level, which is taken to be (as usual) 84.5 percent.

Given the criterion level and the unit size, we can calculate the cumulative distributions for a set of sample sizes. When these are tabulated, there will be a point beyond which the given performance criterion level would produce sample results for, at most, ten percent of the time. That is, the numbers of successes at that point would be exceeded in, at most, ten percent of the samples of the indicated sizes. The limiting number of failures could be graphed versus sample size, or divided by sample size and graphed as failure rate versus sample size.

TABLE 3

TYPICAL SAMPLE SIZES REQUIRED FOR
WEIGHTED SAMPLING (LOWER CRITERION = 84.5 percent)

	True unit rating (percent)					
Unit size	85	86	88	90	95	100
100	100	100	100	95	48	12
200	200	200	196	157	56	12
300	300	300	280	193	60	12
500	500	500	404	232	62	12
1000	1000	991	562	268	64	12
2000	2000	1817	676	288	65	12
3000	3000	2391	722	296	66	12
4000	4000	2790	746	300	66	12
5000	5000	3079	761	302	66	12
6000	5996	3296	771	304	66	12
7000	6987	3465	779	305	66	12
Without replacement	44278	4784	826	312	67	13

TABLE 4

TYPICAL SAVINGS IN SAMPLE SIZES BY USING WEIGHTED SAMPLING (LOWER CRITERION = 84.5 percent)

Unit size	True unit rating (percent) 85	86	88	90	95	100
100	0	0	0	5	19	1
200	0	0	4	43	11	1
300	0	0	20	107	7	1
500	0	0	96	80	5	1
1000	0	9	264	44	3	1
2000	0	183	150	24	2	1
3000	0	609	104	16	1	1
4000	0	1210	80	12	1	1
5000	0	1705	65	10	1	1
6000	4	1488	55	8	1	1
7000	13	1319	47	7	1	1

We now know the results that give satisfactory significance levels with given size samples, but not the probability of obtaining those results. To determine this probability, we repeat the previous calculation, using a postulated true unit rating in place of the criterion level. The tabulation now shows success probabilities when the true unit rating has the assumed value. There will be a region in this table such that 90 percent of the samples will be found in the region, for a given sample size. Again the low failure rate end is selected. This table produces another graph of failure rate versus sample size, which is to be superimposed on the first graph. The intersection point represents failure probabilities that, with the indicated sample size, will be bettered 90 percent of the time (given our assumed true unit rating) and would occur, at most, ten percent of the time if the true unit rating were really the criterion value. Repeating this calculation produces the required sample sizes without replacement as a function of unit size, true unit performance rating, significance level, and success probability. Table 5 shows the results.

In practice, the calculation described above would be very tedious for large sample sizes, even though a computer program, described in appendix A, has been written to calculate the terms and make the comparisons. For large sample sizes, a drastic simplification has been found. The normal approximation to the hypergeometric distribution for large unit sizes is given by (reference 3, page 247):

$$\mu = rp,$$

$$\sigma = rp(1-p)\ \frac{n-r}{n-1} \quad ,$$

where μ is the mean number of passes, r is the sample size, p is the pass probability, and n is the unit size. The probability that the number of passes will fall in the range from a to b is given by:

$$\Pr(a \leq X \leq b) = \Phi\left(\frac{b+1/2-\mu}{\sigma}\right) - \Phi\left(\frac{a+1/2-\mu}{\sigma}\right) \quad ,$$

where Φ is the cumulative normal distribution. For the upper 10 percent, for example,

$$1 - \Phi\left(\frac{a+1/2-\mu}{\sigma}\right) = 0.1,$$

and

$$\frac{a-\mu}{\sigma} \simeq 1.28 \quad .$$

TABLE 5

SAMPLE SIZE REQUIREMENTS CALCULATED FROM THE HYPERGEOMETRIC DISTRIBUTION[a]

True unit rating	Unit size: 50	100	200	300	500	1,000	2,000	3,000	4,000	5,000	6,000	20,000
100	10	13	14	14	14	14	14	14	14	14	14	14
99		22	23	23	24	24	24	24	24	24	24	24
98	20	22	23	23	24	24	24	24	24	24	24	24
97		29	32	32	32	33	33	33	33	33	33	33
96	26	29	39	40	41	41	41	42	42	42	42	42
95		36	47	48	49	49	58	58	58	58	58	58
94	32	43	55	55	64	65	66	66	66	66	66	66
93		50	69	70	79	88	89	89	89	89	89	89
92	38	56	76	84	101	110	119	119	119	119	119	120
91		62	96	105	129	146	155	163	163	163	163	171
90	43	73	108	131	163	196	219	227	234	235	235	243
89		78	133	163	216	271	324	332	347	347	355	370
88	47	88	157	200	280	392	487	530	559	567	582	618
87		93	178	241	355	567	785	911	975	1,024	1,059	1,207
86	50	97	193	274	437	791	1,300	1,659	1,924	2,121	2,263	3,118
85		100	200	297	493	974	1,892	2,758	3,584	4,365	5,099	12,529

[a]Criterion value = 84.5 percent
Success probability = 90 percent
Significance level = 0.1

To deal with the distribution of results when the population mean is the criterion value, $\mu=p_c r$ (where p_c is the criterion percentage) and $a=p_o r$ (where p_o is the percentage that defines the lower boundary of the ten-percent tail). Therefore,

$$r(p_o-p_c) = 1.28\ \sigma_c,$$

where σ_c is the standard deviation when p_c is the mean success rate. Similarly, the lower ten-percent tail of the true rating sample distribution is given by:

$$r(p_o-p_T) = -1.28\,\sigma_T\ .$$

Since σ is a function of r, p, and N, the last two equations involve the known terms p_c, p_T, and N, and the unknown terms r and p_o. They are easily solved for the variable of interest, the sample size r. (If either the success probability or the desired confidence level is changed, the constant 1.28 is changed, and the formulas are slightly altered. The program has this flexibility.) Results of this calculation are shown in tables 6 through 9.

Comparing the columns of table 5 with the appropriate columns of tables 6 through 9 shows that the percentage differences are quite small for true unit ratings of 90 and below. Fortunately, this is the very region in which the exact calculations are tedious. Above these ratings, the sample sizes are not only easier to calculate, but also tend to become constant for unit sizes in our range. For unit ratings above 90, we, therefore, have easily replaced the approximate results of tables 6 through 9 with the exact results. Tables 10 through 13 give the results of this hybrid procedure.

TABLE 6

SAMPLE SIZE REQUIREMENTS: NORMAL APPROXIMATION, UNIT SIZE OF 10(10)100

NORM

CRITERION:84.50 PROBABILITY:90 CONF.LEVEL:10 BETA2:1.282 BETA1:1.282

TRUE UNIT RATING	SAMPLE SIZE REQUIRED									
	10	20	30	40	50	60	70	80	90	100
100	5	7	8	8	8	8	9	9	9	9
99	7	10	11	12	13	14	14	14	15	15
98	8	11	14	15	16	17	18	18	19	19
97	8	13	16	18	19	21	22	22	23	24
96	9	14	18	20	23	24	26	27	28	29
95	9	15	20	23	26	28	30	32	33	34
94	9	16	21	26	29	32	35	37	39	40
93	10	17	23	28	32	36	39	42	45	47
92	10	18	25	31	36	40	45	48	52	55
91	10	18	26	33	39	45	50	54	59	63
90	10	19	27	35	42	49	55	61	66	71
89	10	20	28	37	45	52	59	66	73	79
88	10	20	29	38	47	55	64	72	79	87
87	10	20	30	39	49	58	67	76	85	93
86	10	20	30	40	50	60	69	79	88	98
85	10	20	30	40	50	60	70	80	90	100

TABLE 7

SAMPLE SIZE REQUIREMENTS: NORMAL APPROXIMATION, UNIT SIZE OF 100(100)1,000

CRITERION:84.50 PROBABILITY:90 CONF.LEVEL:10 BETA2:1.282 BETA1:1.282

TRUE UNIT RATING	SAMPLE SIZE REQUIRED									
	100	200	300	400	500	600	700	800	900	1000
100	9	9	9	9	9	9	9	9	9	9
99	15	16	16	17	17	17	17	17	17	17
98	19	21	22	22	22	22	23	23	23	23
97	24	27	28	28	29	29	29	29	29	29
96	29	33	35	36	36	37	37	37	38	38
95	34	41	44	45	46	47	47	48	48	48
94	40	50	54	57	58	60	60	61	62	62
93	47	61	68	72	74	76	78	79	80	80
92	55	75	85	91	96	99	101	103	104	105
91	63	91	107	117	124	129	133	136	139	141
90	71	109	133	150	162	171	178	184	189	193
89	79	130	166	193	213	229	243	254	263	271
88	87	153	204	246	280	309	333	354	372	388
87	93	174	244	306	360	409	453	493	530	563
86	98	190	278	361	441	516	589	657	723	787
85	100	199	298	396	493	590	686	782	877	972

TABLE 8

SAMPLE SIZE REQUIREMENT: NORMAL APPROXIMATION, UNIT SIZE OF 1,000(1,000)10,000

CRITERION:84.50 PROBABILITY:90 CONF.LEVEL:10 BETA2:1.282 BETA1:1.282

TRUE UNIT RATING	SAMPLE SIZE REQUIRED									
	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
100	9	9	9	9	9	9	9	9	9	9
99	17	17	17	17	17	17	17	17	17	17
98	23	23	23	23	23	23	23	23	23	23
97	29	30	30	30	30	30	30	30	30	30
96	38	38	39	39	39	39	39	39	39	39
95	48	49	50	50	50	50	50	50	50	50
94	62	64	65	65	65	65	65	65	65	66
93	80	84	85	85	86	86	86	86	86	86
92	105	111	113	114	115	115	116	116	116	116
91	141	152	155	158	159	160	160	161	161	161
90	193	213	221	225	228	229	231	232	232	233
89	271	313	330	339	345	349	352	354	356	357
88	388	481	523	547	562	573	581	587	592	596
87	563	782	899	971	1021	1057	1084	1106	1123	1137
86	787	1295	1652	1915	2117	2278	2409	2517	2608	2686
85	972	1889	2757	3579	4359	5100	5805	6476	7116	7727

TABLE 9

SAMPLE SIZE REQUIREMENT: NORMAL APPROXIMATION, UNIT SIZE OF 11,000(1,000)20,000

CRITERION:84.50 PROBABILITY:90 CONF.LEVEL:10 BETA2:1.282 BETA1:1.282

TRUE UNIT RATING	SAMPLE SIZE REQUIRED									
	11000	12000	13000	14000	15000	16000	17000	18000	19000	20000
100	9	9	9	9	9	9	9	9	9	9
99	17	17	17	17	17	17	17	17	17	17
98	23	23	23	23	23	23	23	23	23	23
97	30	30	30	30	30	30	30	30	30	30
96	39	39	39	39	39	39	39	39	39	39
95	50	50	50	50	50	50	50	50	50	50
94	66	66	66	66	66	66	66	66	66	66
93	86	86	87	87	87	87	87	87	87	87
92	116	117	117	117	117	117	117	117	117	117
91	162	162	162	162	162	162	162	162	163	163
90	234	234	234	235	235	235	235	235	236	236
89	358	359	360	361	361	362	362	363	363	363
88	599	602	604	606	608	609	611	612	613	614
87	1149	1159	1167	1175	1182	1187	1193	1197	1201	1205
86	2753	2812	2863	2909	2950	2986	3020	3050	3077	3102
85	8311	8869	9404	9916	10407	10879	11332	11768	12187	12591

TABLE 10

SAMPLE SIZE REQUIREMENTS: UNIT SIZES OF 10(10)100[a]

TRUE UNIT RATING	UNIT SIZE 10	20	30	40	50	60	70	80	90	100
100	7	10	10	10	10	12	13	13	13	13
99	8	12	15	16	16	16	19	20	22	22
98	8	14	17	20	20	21	21	21	22	22
97	9	14	19	23	24	24	25	25	28	29
96	9	14	21	24	26	26	26	29	29	29
95	9	14	23	24	26	26	31	35	36	36
94	10	17	24	30	32	32	36	40	40	43
93	10	18	25	32	33	34	41	45	46	50
92	10	19	26	34	36	36	45	49	52	56
91	10	19	27	35	39	40	48	54	62	62
90	10	19	27	35	42	49	55	61	66	71
89	10	20	28	37	45	52	59	66	73	79
88	10	20	29	38	47	55	64	72	79	87
87	10	20	30	39	49	58	67	76	85	93
86	10	20	30	40	50	60	69	79	88	98
85	10	20	30	40	50	60	70	80	90	100

[a]Criterion value = 84.5 percent
Success probability = 90 percent
Significance level = 0.1

TABLE 11

SAMPLE SIZE REQUIREMENTS: UNIT SIZES OF 100(100)1,000[a]

TRUE UNIT RATING	UNIT SIZE									
	100	200	300	400	500	600	700	800	900	1000
100	13	14	14	14	14	14	14	14	14	14
99	22	23	23	24	24	24	24	24	24	24
98	22	23	23	24	24	24	24	24	24	24
97	29	32	32	32	32	33	33	33	33	33
96	29	39	40	41	41	41	41	41	41	41
95	36	47	48	49	49	49	49	49	49	49
94	43	55	55	64	64	65	65	65	65	65
93	50	69	70	79	79	80	80	80	88	88
92	56	76	91	93	101	102	102	110	110	110
91	62	96	111	121	129	130	137	139	146	146
90	71	109	133	150	162	171	178	184	189	193
89	79	130	166	193	213	229	243	254	263	271
88	87	153	204	246	280	309	333	354	372	388
87	93	174	244	306	360	409	453	493	530	563
86	98	190	278	361	441	516	589	657	723	787
85	100	199	298	396	493	590	686	782	877	972

[a]Criterion value = 84.5 percent
Success probability = 90 percent
Significance level = 0.1

TABLE 12

SAMPLE SIZE REQUIREMENTS: UNIT SIZES OF 1,000(1,000)10,000[a]

TRUE UNIT RATING	UNIT SIZE 1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
100	14	14	14	14	14	14	14	14	14	14
99	24	24	24	24	24	24	24	24	24	24
98	24	24	24	24	24	24	24	24	24	24
97	33	33	33	33	33	33	33	33	33	33
96	41	41	42	42	42	42	42	42	42	42
95	49	58	58	58	58	58	58	58	58	58
94	65	66	66	66	66	66	66	66	66	66
93	88	89	89	89	89	89	89	89	89	89
92	110	119	119	119	119	119	119	119	119	119
91	146	155	163	163	163	163	163	164	164	165
90	193	213	221	225	228	229	231	232	232	233
89	271	313	330	339	345	349	352	354	356	357
88	388	481	523	547	562	573	581	587	592	596
87	563	782	899	971	1021	1057	1084	1106	1123	1137
86	787	1295	1652	1915	2117	2278	2409	2517	2608	2686
85	972	1889	2757	3579	4359	5100	5805	6476	7116	7727

[a]Criterion value = 84.5 percent
Success probability = 90 percent
Significance level = 0.1

TABLE 13

SAMPLE SIZE REQUIREMENTS: UNIT SIZES OF 11,000(1,000)20,000[a]

TRUE UNIT RATING	UNIT SIZE									
	11000	12000	13000	14000	15000	16000	17000	18000	19000	20000
100	14	14	14	14	14	14	14	14	14	14
99	24	24	24	24	24	24	24	24	24	24
98	24	24	24	24	24	24	24	24	24	24
97	33	33	33	33	33	33	33	33	33	33
96	42	42	42	42	42	42	42	42	42	42
95	58	58	58	58	58	58	58	58	58	58
94	66	66	66	66	66	66	66	66	66	66
93	89	89	89	89	89	89	89	89	89	89
92	120	120	120	120	120	120	120	120	120	120
91	165	166	166	167	167	168	169	169	170	170
90	234	234	234	235	235	235	235	235	236	236
89	358	359	360	361	361	362	362	363	363	363
88	599	602	604	606	608	609	611	612	613	614
87	1149	1159	1167	1175	1182	1187	1193	1197	1201	1205
86	2753	2812	2863	2909	2950	2986	3020	3050	3077	3102
85	8311	8869	9404	9916	10407	10879	11332	11768	12187	12591

[a]Criterion value = 84.5 percent
Success probability = 90 percent
Significance level = 0.1

REFERENCES

1. Sokal, R., and Rohlf, F., "Biometry," W.H. Freeman and Company, San Francisco, 1969, LC 68-16819

2. Feller, W., "An Introduction to Probability Theory and Its Applications," Volume I, Second Edition, John Wiley & Sons, Inc., New York, 1960, LC 57-10805

3. Meyer, S., "Data Analysis for Scientists and Engineers," John Wiley & Sons, Inc., New York, 1975, LC 74-8873

4. Bailey, D.E., "Probability and Statistics: Models for Research," John Wiley & Sons, Inc., New York, 1971, LC 75-130428

APPENDIX A

COMPUTER PROGRAMS

APPENDIX A

COMPUTER PROGRAMS

This appendix presents and discusses the computer programs for the following calculations:

- Sample sizes required for exceeding a criterion level (with replacement);
- Sample sizes required for staying below a criterion level (with replacement);
- Sample sizes required for staying inside a pair of criterion levels (with replacement);
- Sample size requirements from the hypergeometric distribution;
- Sample size requirements from the normal approximation to the hypergeometric distribution;
- Printing the final size requirement matrixes.

All programs are written in APL/700 for a Burroughs B6000/B7000 computer.

Two-Level Calculation (With Replacement)

Table A-1 shows the program ABOVE, which calculates the sample size required for a given probability P that a sample rating will be above a given criterion level, at some given significance level α , given the true unit rating. The success probabilities included are .5, .8, .85, .9, .95, and .99. The significance levels are .1, .05, .01, and .001 . The criterion level and maximum unit rating are input as percentages. A sample output is also shown in table A-1.

The rationale is from reference 1 (pages 607-610), where it is said to be inappropriate for "very small" sample sizes. The sample size estimate is based on a t-test for the equality of two percentages, where t is given by:

$$t = \frac{\text{Arcsin}\sqrt{p_1} - \text{Arcsin}\sqrt{p_2}}{\sqrt{820.8\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} .$$

Here p_i is a fraction, N_i is a sample size, and the angles are in degrees.

N_2 is set "equal" to infinity to represent the criterion value. The t-value for the assumed probability of success and the value for the assumed significance level are then averaged to obtain an effective t (reference 4, page 349). The minimum sample size requirement is then given by:

TABLE A-1

PROGRAM AND SAMPLE OUTPUT: CALCULATION OF SAMPLE SIZE REQUIRED (WITH REPLACEMENT) TO OBTAIN RESULTS ABOVE A CRITERION LEVEL

```
     ∇ABOVE[⎕]∇
    ∇ ABOVE
[1]   I←(PROB=0.5)+(2×PROB=0.8)+(3×PROB=0.85)+(4×PROB=0.9)+(
      5×PROB=0.95)+6×PROB=0.99
[2]   J←(A=0.1)+(2×A=0.05)+(3×A=0.01)+4×A=0.001
[3]   'PROBABILITY OF SUCCESS: ';4 2⍕PROB
[4]   'SIGNIFICANCE LEVEL: ';5 3⍕A
[5]   P←((P1-⌊P2),2)⍴0
[6]   P[;1]←((⌊P2)+⍳P1-⌊P2)×0.01
[7]   P[;2]←((P1-⌊P2)⍴P2)×0.01
[8]   N←0.5×(CONST[I;J]÷(180÷○1)⋆2)÷((¯1○P[;1]⋆0.5)-¯1○P[;2]
      ⋆0.5)⋆2
[9]   NN←((⍴N),2)⍴0
[10]  NN[;1]←P[;1]
[11]  NN[;2]←⌊N+0.5
[12]  'LOWER CUTOFF PERCENTAGE: ';4 1⍕P2
[13]  ' '
[14]  '    ACTUAL  SAMPLE'
[15]  '     UNIT    SIZE'
[16]  '    RATING REQUIRED'
[17]  '    ------ --------'
[18]  NN
    ∇
```

```
              PROBABILITY OF SUCCESS: 0.95
              SIGNIFICANCE LEVEL: 0.050
              LOWER CUTOFF PERCENTAGE: 84.5

                 ACTUAL  SAMPLE
                  UNIT    SIZE
                 RATING REQUIRED
                 ------ --------
                  0.85   67186
                  0.86    7259
                  0.87    2537
                  0.88    1254
                  0.89     733
                  0.9      473
                  0.91     325
                  0.92     234
                  0.93     173
                  0.94     132
                  0.95     101
                  0.96      79
                  0.97      61
                  0.98      47
                  0.99      35
                  1         20
```

$$N = \frac{2x820.8\ \{t_\alpha + t_{2(1-P)}\}^2}{\{\text{Arcsin}\sqrt{p_1} - \text{Arcsin}\sqrt{p_2}\}^2} .$$

The numerator is not a function of the ratings p_i; it has been tabulated for a range of α and P. Table A-2 shows most of the values derivable from standard tables of the t-distribution.

(Most of the data in table A-2 comes from page 609 of reference 1, where the last two columns are mislabeled.)

TABLE A-2

COEFFICIENTS FOR THE SAMPLE SIZE CALCULATIONS

PROB. OF SUCCESS	CONFIDENCE LEVEL 0.1	0.05	0.01	0.001
0.5	4442.2	6306.4	10891.5	17780.0
0.8	10150.2	12884.8	19171.6	28036.0
0.85	11841.0	14781.0	21473.0	30802.0
0.9	14059.3	17249.8	24426.2	34324.0
0.95	17768.0	21333.0	29247.0	39994.0
0.99	25890.0	30161.4	39450.1	51794.0

Table A-3 shows the program and output when the requirement is to show that a rating is below a criterion level. In both cases (ABOVE and BELOW), the parameters to be set are the success probability desired (PROB, from the list given above), the confidence level (A, also from the above list), and the cutoff percentages (P2 if lower and P1 if upper). The matrix CONST, given as table A-2, is also required.

Three-Level Calculations (With Replacement)

When there are three or more levels, units can be confidently placed in intermediate levels by establishing that the true unit rating is likely to be simultaneously above a lower cutoff and below an upper cutoff. Table A-4 gives a program for this calculation, and table A-5 shows part of a typical output. The rationale is described in the main body of this report. Some program details bear discussion here.

TABLE A-3

PROGRAM AND SAMPLE OUTPUT: CALCULATION OF SAMPLE SIZE REQUIRED (WITH REPLACEMENT) TO OBTAIN RESULTS BELOW A CRITERION LEVEL

```
      ∇BELOW[⎕]∇
    ∇ BELOW
[1]   I←(PROB=0.5)+(2×PROB=0.8)+(3×PROB=0.85)+(4×PROB=0.9)+(
      5×PROB=0.95)+6×PROB=0.99
[2]   J←(A=0.1)+(2×A=0.05)+(3×A=0.01)+4×A=0.001
[3]   'PROBABILITY OF SUCCESS: ';4 2⍕PROB
[4]   'SIGNIFICANCE LEVEL: ';5 3⍕A
[5]   P←(((⌈P1)-P2+1),2)⍴0
[6]   P[;1]←(P2+⍳(⌈P1)-P2+1)×0.01
[7]   P[;2]←(((⌈P1)-P2+1)⍴P1)×0.01
[8]   N←0.5×(CONST[I;J]÷(180÷○1)*2)÷((¯1○P[;1]*0.5)-¯1○P[;2]
      *0.5)*2
[9]   NN←((⍴N),2)⍴0
[10]  NN[;1]←P[;1]
[11]  NN[;2]←⌊N+0.5
[12]  'UPPER CUTOFF PERCENTAGE: ';4 1⍕P1
[13]  ' '
[14]  '    ACTUAL  SAMPLE'
[15]  '     UNIT    SIZE'
[16]  '    RATING REQUIRED'
[17]  '    ------ --------'
[18]  NN
    ∇
```

```
PROBABILITY OF SUCCESS: 0.95
SIGNIFICANCE LEVEL: 0.050
UPPER CUTOFF PERCENTAGE: 97.0

  ACTUAL  SAMPLE
   UNIT    SIZE
  RATING REQUIRED
  ------ --------

  0.86      74
  0.87      86
  0.88     101
  0.89     121
  0.9      149
  0.91     190
  0.92     256
  0.93     370
  0.94     603
  0.95    1228
  0.96    4368
```

TABLE A-4

PROGRAM FOR CALCULATION OF
SAMPLE SIZE REQUIRED
(WITH REPLACEMENT) TO OBTAIN RESULTS
WITHIN A RATING LEVEL

```
      ∇INSIDE[⎕]∇
     ∇ INSIDE
[1]    'UPPER LIMIT: ';P1
[2]    'LOWER LIMIT: ';P2
[3]    'SIGNIFICANCE LEVEL: 0.1'
[4]    P←P2+1
[5]   AGAIN:II←1
[6]    M←(ρCC[;1])ρ0
[7]   LOOP:M[II]←CC[II;2]÷((¯1○(P×0.01)*0.5)-¯1○(P2×0.01)*
       0.5)*2
[8]    II←II+1
[9]    →LOOP×⍳II≤ρM
[10]   DELTA←|((¯1○(P×0.01)*0.5)-¯1○(P1×0.01)*0.5)×180÷○1
[11]   T←(DELTA×(M÷820.8)*0.5)-1.645
[12]   T←(T×T≤2.576)+2.576×T>2.576
[13]   J←1
[14]   PP←(ρM)ρ0
[15]   →SWITCH×⍳+/T<0.542
[16]  JJ:PP[J]←+/0.4954 0.41581 ¯0.01178 ¯0.084859 0.0297
       ¯0.0030905×T[J]*0 1 2 3 4 5
[17]   J←J+1
[18]   →JJ×⍳J<1+ρM
[19]   →SET
[20]  SWITCH:II←1
[21]  LOOP2:M[II]←CC[II;2]÷((¯1○(P×0.01)*0.5)-¯1○(P1×0.01)*
       0.5)*2
[22]   II←II+1
[23]   →LOOP2×⍳II≤ρM
[24]   DELTA←|((¯1○(P×0.01)*0.5)-¯1○(P2×0.01)*0.5)×180÷○1
[25]   T←(DELTA×(M÷820.8)*0.5)-1.645
[26]   T←(T×T≤2.576)+2.576×T>2.576
[27]   J←1
[28]  JJ1:PP[J]←+/0.4954 0.41581 ¯0.01178 ¯0.084859 0.0297
       ¯0.0030905×T[J]*0 1 2 3 4 5
[29]   J←J+1
[30]   →JJ1×⍳J<1+ρM
[31]   MM←((ρM),4)ρ0
[32]   MM[;2]←CC[;1]
[33]   MM[;1]←PP×T≥0.542
[34]   MM[;3]←(MM[;1]+MM[;2]-1)×T≥0.542
[35]   MM[;4]←⌊M+0.5
```

TABLE A-4 (CONT'D)

```
[36]  →PRINT
[37] SET:MM←((ρM),4)ρ0
[38]  MM[;1]←CC[;1]
[39]  MM[;2]←PP×T≥0.542
[40]  MM[;3]←(MM[;1]+MM[;2]-1)×T≥0.542
[41]  MM[;4]←LM+0.5
[42] PRINT:'~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~'
[43]  ' '
[44]  'TRUE RATING: ';P
[45]  ' '
[46]  '              PROBABILITIES          NUMBER OF'
[47]  '              -------------             OF  '
[48]  '    LOWER     UPPER     INSIDE       CASES'
[49]  '    -----     -----     ------       -----'
[50]  K←1
[51] DO:9 4⍕MM[K;1];9 4⍕MM[K;2];9 4⍕MM[K;3];9 0⍕MM[K;4]
[52]  K←K+1
[53]  →DO×⍳K≤ρM
[54]  ' '
[55]  PROB←80
[56] OUT:S←(ρM)+1-+/MM[;3]<PROB×0.01
[57]  L←S-1
[58]  NUMBER←10*(10⍟MM[S;4])+(10⍟MM[L;4]÷MM[S;4])×(10⍟(PROB×
      0.01)÷MM[S;3])÷10⍟MM[L;3]÷MM[S;3]
[59]  'NUMBER OF CASES FOR ';PROB;' PERCENT PROBABILITY IN
      REGION: ';5 0⍕⌊NUMBER+0.5
[60]  PROB←PROB+5
[61]  →OUT×⍳PROB<100
[62]  →NEW×⍳PROB=104
[63]  PROB←99
[64]  →OUT
[65] NEW:P←P+1
[66]  →AGAIN×⍳P<P1
[67]  →0
```

It is assumed that no probabilities exceed 99.5 percent. (In principle, this is easily extended to 99.9 percent.) When any probability is less than 70 percent ($t < 0.542$), the program switches from calculating the sample size initially for the lower criterion to calculating it initially for the upper criterion (step 15). This keeps the composite probabilities within the range of interest. The example output in table A-5 shows this shift. When the composite probabilities have been calculated for each of seven one-level probabilities (ranging from .7 to .995), logarithmic interpolation is used to obtain the sample size requirements for probabilities of 80, 85, 90, and 95 percent. The program is written for a 0.1 significance level, which is indicated in step 3 and determined by the constant $t_\alpha = 1.645$ in steps 11 and 25. Generalization to a set of significance levels is straightforward.

The input to program INSIDE is nearly the same as for ABOVE and BELOW, except that the constant is now given by the matrix CC, which includes the factor 2 and the conversion from radians to degrees. This matrix is given in table A-6.

Sample Size Requirements from the Hypergeometric Distribution

Three nested programs were used to calculate sample sizes from hypergeometric distributions. Table A-7 shows the main routine CALC. Table A-8 shows the subroutines. HYPRR calculates the cumulative hypergeometric distributions; STIR is a subroutine of HYPRR, used to calculate the logarithms of binomial coefficients.

The calculation presents two problems: numbers that are too large or too small to handle, and matrixes that are too large for an APL workspace. Even when the distribution terms are not small enough to neglect, the binomial coefficients required may be out of the computer range. STIR avoids this problem by calculating the logarithms of the coefficients. STIR is slightly complicated by the necessity to handle binomial coefficients that use zero or negative factorials. The routine uses a modified Stirling's formula (reference 2, page 52).

HYPRR calculates distribution terms in two ways. First, it simply calculates the terms, using STIR, until a term exceeding some tolerance level, TOL, is found. From that point on, the terms are calculated by means of the following recursion relation (reference 3, page 194):

$$H\{I,J\} = \frac{(K+2-I)\ (J+2-I)}{(I-1)\ (N+I-K-J)} \quad H\{I-1,J\} ,$$

TABLE A-5

PARTIAL PROGRAM OUTPUT FOR CALCULATION
OF SAMPLE SIZE REQUIRED (WITH REPLACEMENT) TO
OBTAIN RESULTS WITHIN A RATING LEVEL

TRUE RATING: 95.5

PROBABILITIES			NUMBER OF
LOWER	UPPER	INSIDE	CASES
0.[illegible]50	0.9950	0.9900	122
0.9900	0.9950	0.9850	108
0.9750	0.9915	0.9665	89
0.9500	0.9793	0.9293	74
0.9000	0.9487	0.8487	59
0.8000	0.8725	0.6725	42
0.7000	0.7894	0.4894	33

NUMBER OF CASES FOR 80 PERCENT PROBABILITY IN REGION: 54
NUMBER OF CASES FOR 85 PERCENT PROBABILITY IN REGION: 59
NUMBER OF CASES FOR 90 PERCENT PROBABILITY IN REGION: 68
NUMBER OF CASES FOR 95 PERCENT PROBABILITY IN REGION: 82

TRUE RATING: 96.5

PROBABILITIES			NUMBER OF
LOWER	UPPER	INSIDE	CASES
0.9950	0.9950	0.9900	126
0.9950	0.9900	0.9850	111
0.9938	0.9750	0.9688	92
0.9838	0.9500	0.9338	76
0.9575	0.9000	0.8575	60
0.8876	0.8000	0.6876	44
0.8080	0.7000	0.5080	34

NUMBER OF CASES FOR 80 PERCENT PROBABILITY IN REGION: 54
NUMBER OF CASES FOR 85 PERCENT PROBABILITY IN REGION: 59
NUMBER OF CASES FOR 90 PERCENT PROBABILITY IN REGION: 69
NUMBER OF CASES FOR 95 PERCENT PROBABILITY IN REGION: 83

TABLE A-6

CONSTANTS MATRIX (CC) FOR PROGRAM "INSIDE"

Probability P	Constant[a]
.995	4.454
.99	3.942
.975	3.249
.95	2.706
.9	2.142
.8	1.546
.7	1.196

---

[a]Constant = 2 x $(180/\pi)^2$ x 820.8 x

$$\left[t_\alpha + t_{2(1-P)} \right]^2 ,$$

where $\alpha = 0.1$.

TABLE A-7

ROUTINE FOR CALCULATING
SAMPLE SIZES FROM
HYPERGEOMETRIC DISTRIBUTION

```
      ∇CALC[⎕]∇
     ∇ CALC
[1]    K2←K20
[2]   NEWJ:J2←J1+DJ
[3]    E←J1×K1÷N
[4]    SIGMA←(E×(N-J1)×(N-K1)÷N*2)*0.5
[5]    ROW1←1⌈+⌊E+SIGMA
[6]    E1←J2×K2÷N
[7]    SIGMA1←(E1×(N-J2)×(N-K2)÷N*2)*0.5
[8]    ROW2←⌈(⌈E1-SIGMA1)⌈E+X2×SIGMA
[9]    ROW1←⌈1⌈ROW1⌊E1-X1×SIGMA1
[10]   K←K1
[11]   HM1←HYPRR
[12]   →NEWK×⍳SW=0
[13]   PICK←((⌊/HM1)≤0.2)∧(⌈/HM1)≥0.001
[14]   PICK←PICK/⍳1+ROW2-ROW1
[15]   ((7+2×DJ≤4),0)⍕(XX≠J1-1)×XX←¯2+J1+⍳2+J2-J1
[16]   ' '
[17]   ((7+2×DJ≤4),2)⍕(⍉(¯2+ROW1)+⍳1+ROW2-ROW1)[PICK;],[2]
       HM1[PICK;]
[18]   ' '
[19]   ' '
[20]  NEWK:K←K2
[21]   HM2←HYPRR
[22]   →SKIP×⍳SW=0
[23]   PICK←((⌊/HM2)≤0.2)∧(⌈/HM2)≥0.001
[24]   PICK←PICK/⍳1+ROW2-ROW1
[25]   ((7+2×DJ≤4),0)⍕(XX≠J1-1)×XX←¯2+J1+⍳2+J2-J1
[26]   ' '
[27]   ((7+2×DJ≤4),2)⍕(⍉(¯2+ROW1)+⍳1+ROW2-ROW1)[PICK;],[2]
       HM2[PICK;]
```

```
[28] SKIP:' '
[29]  HM1A←HM1≤0.1
[30]  HM2A←HM2≤0.1
[31]  H1←1+(⍴HM1A[;1])-+/HM1A
[32]  H2←1++/HM2A
[33]  H←H1-H2
[34]  HH←((1+J2-J1),2)⍴0
[35]  HH[;1]←(J1-1)+⍳1+J2-J1
[36]  HH[;2]←H
[37]  'UNIT SIZE: ';N
[38]  'PASSES(CRITERION): ';K1;'(';5 1⍕100×K1÷N;' PERCENT)'
[39]  'PASSES(TRUE RATING): ';K2;'(';5 1⍕100×K2÷N;' PERCENT)
      '
[40]  ' '
[41]  5 0⍕HH
[42]  '~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      '
[43]  K2←K2-DP
[44]  →NEWK×⍳K2>K1
[45]  J1←J1+DJ+1
[46]  K2←K20
[47]  '|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      |||||||||||||||||||||||||||||||||||||'
[48]  →NEWJ×⍳J2≤LIMIT-DJ+1
    ∇
```

TABLE A-8

SUBROUTINES FOR CALCULATING SAMPLE
SIZE REQUIREMENTS FROM THE HYPERGEOMETRIC DISTRIBUTION

```
     ∇HYPRR[⎕]∇
    ∇ HHH←HYPRR
[1]    HHH←((1+ROW2-ROW1),1+J2-J1)⍴0
[2]    J←1
[3]    I←1
[4]   COL1:→COL11×⍳(J1+J+1-ROW1+I)≤N-K
[5]    HHH[I;J]←0
[6]    →II
[7]   COL11:HHH[I;J]←*((ROW1+I-2)STIR K)+((J1+J+1-ROW1+I)
       STIR N-K)-(J1+J-1)STIR N
[8]    →RECURS×⍳HHH[I;J]≥TOL
[9]   II:I←I+1
[10]   →COL1×⍳(I≤J1+J+1-ROW1)∧I≤1+ROW2-ROW1
[11]  JJ:J←J+1
[12]   I←1
[13]   →COL1×⍳J≤1+J2-J1
[14]   →SUM
[15]  RECURS:I←I+1
[16]   →JJ×⍳I>1+ROW2-ROW1
[17]   HHH[I;J]←HHH[I-1;J]×(K+3-I+ROW1)×(J1+J+2-I+ROW1)÷(I+
       ROW1-2)×N+I+ROW1-K+J+J1+1
[18]   →ON×⍳HHH[I;J]≥HHH[I-1;J]
[19]   →JJ×⍳HHH[I;J]<1E¯6
[20]  ON:→RECURS×⍳I<1+ROW2-ROW1
[21]   →JJ
[22]  SUM:→SUM1×⍳K=K2
[23]   HHH←⊖+\⊖HHH
[24]   →0
[25]  SUM1:HHH←+\HHH
[26]   →0
    ∇
```

TABLE A-8 (CONT'D)

```
      ∇STIR[⎕]∇
    ∇ LN←X STIR Z
[1]    →EQUAL×⍳Z=X
[2]    →ZERO1×⍳Z=0
[3]    LN1←FACT+((Z+0.5)×⍟Z)+(÷12×Z)−Z
[4]    →ZERO1+1
[5]   ZERO1:LN1←⍟1
[6]    →ZERO2×⍳X=0
[7]    LN2←FACT+((X+0.5)×⍟X)+(÷12×X)−X
[8]    →ZERO2+1
[9]   ZERO2:LN2←⍟1
[10]   LN3←FACT+((0.5+Z−X)×⍟1⌈Z−X)+(÷12×Z−X)+X−Z
[11]   LN←LN1−LN2+LN3
[12]   LN←LN−30×Z<X
[13]   →0×⍳Z≠X
[14]  EQUAL:LN←0
    ∇
```

where I is one more than the number of successes in the sample, K is the number of successes in the population, J is the sample size, and N is the population size. When a term less than some tolerance level (and also less than the previous term) is reached, the remaining terms are assumed to be zero.

HYPRR and STIR thus allow, in a reasonably efficient manner, the calculation of any distribution terms of practical interest; but, for large units, the number of terms exceeds the workspace capacity. While it is desirable to have a routine that can calculate and sum all terms (to verify directly that the calculated probabilities do sum to one), only a small fraction of the terms are required for the principal purpose.

Consider the case of sample sizes for testing whether a criterion has been exceeded. Only one point on each of two cumulative distributions need be determined. Assuming that the true unit rating is the criterion level, we calculate the point at which the ten-percent upper tail starts. Sample ratings above that point have, at most, a ten-percent chance of being due to true unit ratings at or below the criterion level. We then locate the upper end of the ten-percent lower tail of the distribution with some other, higher, true unit rating assumed. Sample ratings above that point will occur 90 percent of the time if the true unit rating is that assumed. Initially, for small sample sizes, the two tails will not overlap, and the desired success probability and statistical confidence level are not achievable simultaneously. When the two end points merge (or cross), the required sample size has been found. The program is set to indicate that point by displaying the number of successes at the upper ten-percent point of the criterion distribution, less the number at the lower ten-percent point of the true-rating distribution. This number, initially positive, goes to zero at the smallest acceptable sample size and, finally, becomes more and more negative. The initial zero may be followed by positive terms (always ones), representing tradeoffs where a larger sample size produces, for example, a success probability greater than 90 percent, but a confidence level slightly greater than the desired ten-percent. We define the required sample size as that producing the initial zero.

It will be observed that all that is required is to identify the initial zero, which occurs at ten-percent points on each distribution. Because cumulative distributions are required for comparisons, it is necessary to calculate the ten-percent tail terms of each of two hypergeometric distributions. The tails can be very long, however, consuming computer time and space. We calculate the standard deviations of the criterion distribution (SIGMA) and the true-rating distribution (SIGMA1), and calculate only terms within X2 SIGMAs of the true-rating distribution and X1 SIGMA1s of the criterion distribution. By occasionally monitoring the detailed output (the cumulative distribution terms) and the input parameters, the terms calculated can be held to only a few more than absolutely required. The program PAR, table A-9, prints the input parameters to assist in this adjustment. (Initially, a more elegant approach, not requiring user intervention, was developed, but it proved to require a large fraction of the program running time.)

TABLE A-9

PROGRAM FOR PRINTING INPUT PARAMETERS
FOR CALC, PLUS SAMPLE OUTPUT

```
      ∇PAR[⎕]∇
     ∇ PAR
[1]    'PARAMETERS'
[2]    '----------'
[3]    'N= ';6 0⍕N;'      ';'J1= ';J1
[4]    'K1= ';5 0⍕K1;'     ';'K20= ';K20
[5]    'DJ= ';5 0⍕DJ;'     ';'DP= ';DP
[6]    'SW= ';5 0⍕SW
[7]    'LIMIT= ';LIMIT
[8]    'X2= ';2 0⍕X2;'           ';'X1= ';2 0⍕X1
[9]    'SIGMA= ';5 1⍕SIGMA;'    ';'SIGMA1= ';5 1⍕SIGMA1
[10]   'ROW1= ';5 0⍕ROW1;'    ';'ROW2= ';5 0⍕ROW2
     ∇


      PAR
PARAMETERS
----------
N=     100    J1= 60
K1=     76   K20= 82
DJ=      9    DP= 1
SW=      0
LIMIT= 90
X2= 10           X1= 10
SIGMA=    1.7   SIGMA1=    1.2
ROW1=     43   ROW2=     60
```

The programs described above work for unit sizes up to at least 20,000, using piecemeal calculations. This means testing five or ten different sample sizes at a given time. For a unit size as small as ten, the program could simply be run to calculate and print out all terms. With larger unit sizes, this process leads to overlapping lines in the printout, and, fairly quickly, it will also overflow the available computer workspace. Consider the total output for a given unit size. It involves a matrix for the criterion value that has a column for each sample size, and a similar matrix for each true unit rating considered. To determine the sample sizes for ten different true ratings, we need to identify the ten pairs of matrix elements representing the 10-percent points crossovers. With a unit size of 1,000, there are 1,001 elements in a column, 1,000 different sample sizes (rows), and, say, ten true-rating matrixes for approximately eleven million elements, of which only 20 are of direct interest. Many of these elements have already been eliminated by the cutoffs already described, but these grow less useful as more sample sizes are considered at one time. Still, there are many terms to examine to determine a large sample size for a large unit, and, even with efficient selection of regions for examination, the process becomes tedious.

The final step that makes the calculations reasonable is to incorporate a normal approximation into the system (reference 3, page 247). A program for this calculation is given as table A-10, and a partial output is shown as table A-11. When these results were compared to the exact results, it was found that, in the regions where the difference was greater than about two percent, the exact calculations were easy to do. In the other regions, the normal approximation tended to underestimate slightly the required sample sizes.

The approximation produces a full set of sample sizes over a range of unit sizes from 10 to 20,000 (in reasonable increments) in a trivial amount of computer time. These answers can then be used in two ways. First, the easy calculations (small units or true ratings well away from the criterion) are done until the exact sample sizes satisfactorily match the approximate sample sizes, at which point the remainder of the required sample sizes for that unit size are taken from the approximation. Secondly, in this process, the search for the exact size can be started at sizes only a little below the approximate size, adding greatly to the efficiency of the search. Finally, program P (table A-12) is used to print the corrected sample size results in the format of table A-11.

TABLE A-10

PROGRAM TO CALCULATE SAMPLE SIZES FROM NORMAL
APPROXIMATION TO THE HYPERGEOMETRIC DISTRIBUTION

```
      ∇NORM[⎕]∇
    ∇ NORM
[1]   N←10 4⍴0
[2]   N[;1]←10×⍳10
[3]   N[;2]←100×⍳10
[4]   N[;3]←1000×⍳10
[5]   N[;4]←10000+1000×⍳10
[6]   PT←J←K←I←1
[7]  MORE:R←((⌈(PT−PC)×100),⍴N[;K])⍴0
[8]  START:ROOTS←((BETA1*2)×PC×1−PC)+((BETA2*2)×PT×1−PT)+2×BETA1×
      BETA2×(PC×PT×(1−PC)×1−PT)*0.5
[9]   ALPHA←((÷PT−PC)*2)×ROOTS
[10]  R[I;J]←⌈(ALPHA×N[J;K])÷ALPHA+N[J;K]−1
[11]  PT←PT−0.01
[12]  I←I+1
[13]  →START×⍳PT>PC
[14]  J←J+1
[15]  I←PT←1
[16]  →START×⍳J≤⍴N[;K]
[17]  ' '
[18]  ' '
[19]  '   CRITERION:';5 2⍕PC×100;'    ';'PROBABILITY:';2 0⍕PROB;'    ';'
      CONF.LEVEL:';2 0⍕CONF;'    ';'BETA2:';5 3⍕BETA2;'    ';'BETA1:';5
      3⍕BETA1
[20]  ' '
[21]  '     TRUE    '
[22]  '     UNIT                                   SAMPLE SIZE REQUIRED'
                                                   --------------------
[23]  '    RATING                                                        '
[24]  '    ------      '
[25]  '          ';7 0⍕N[;K]
[26]  '          ---------  -----  -----  -----  -----  -----  -----
      -----  -----  -----'
[27]  Q←⌈(1−PC)×100
[28]  R←(⍉101−⍳Q),[2]R
[29]  7 0⍕R
[30]  '~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      ~~~~~~~~~~~~~~~~~'
[31]  K←K+1
[32]  I←PT←J←1
[33]  →MORE×⍳K≤4
[34]  →0
    ∇
```

TABLE A-11

SAMPLE OUTPUT FOR SAMPLE SIZES FROM NORMAL APPROXIMATION
TO THE HYPERGEOMETRIC DISTRIBUTION

CRITERION:84.50 PROBABILITY:90 CONF.LEVEL:10 BETA2:1.282 BETA1:1.282

TRUE UNIT RATING	SAMPLE SIZE REQUIRED									
	100	200	300	400	500	600	700	800	900	1000
100	9	9	9	9	9	9	9	9	9	9
99	15	16	16	17	17	17	17	17	17	17
98	19	21	22	22	22	22	23	23	23	23
97	24	27	28	28	29	29	29	29	29	29
96	29	33	35	36	36	37	37	37	38	38
95	34	41	44	45	46	47	47	48	48	48
94	40	50	54	57	58	60	60	61	62	62
93	47	61	68	72	74	76	78	79	80	80
92	55	75	85	91	96	99	101	103	104	105
91	63	91	107	117	124	129	133	136	139	141
90	71	109	133	150	162	171	178	184	189	193
89	79	130	166	173	213	229	243	254	263	271
88	87	153	204	246	280	309	333	354	372	388
87	93	174	244	306	360	409	453	493	530	563
86	98	190	278	361	441	516	589	657	723	787
85	100	199	298	396	493	590	686	782	877	972

TABLE A-12

PROGRAM FOR PRINTING FINAL SAMPLE SIZE MATRIXES

```
      )SAVE
SAVED 78/12/19 14.56.19 (LJGRIKE)H
      ∇P[⎕]∇
     ∇ P
[1]     '     TRUE'
[2]     '     UNIT                              SAMPLE SIZE REQUIRED'
[3]     '    RATING                             --------------------'
[4]     '    -------'
[5]     '        ';7 0⍕N[;K]
[6]     →((K=1),(K=2),(K=3),K=4)/K1,K2,K3,K4
[7]   K1:'           --     --     --     --     --     --     --     --     --    ---'
[8]     →KK
[9]   K2:'          ---    ---    ---    ---    ---    ---    ---    ---    ---   ----'
[10]    →KK
[11]  K3:'         ----   ----   ----   ----   ----   ----   ----   ----   ----  -----'
[12]    →KK
[13]  K4:'        -----  -----  -----  -----  -----  -----  -----  -----  -----  -----'
[14]  KK:7 0⍕R
     ∇
```

APPENDIX B

A USER'S GUIDE

APPENDIX B

A USER'S GUIDE

This appendix presents some guidance for users of the programs described in appendix A. Appendix A contains the programs and general descriptions and should be read first.

PROGRAMS ABOVE AND BELOW

Select the success probability desired from the following list: .5, .8, .85, .9, .95, .99. (Example : PROB ← .5)

Select the confidence level desired from the following list: .1, .05, .01, .001. (Example: A ← .1)

Specify the criterion value as a percentage. (Example: P2 ← 84.5. For BELOW, use P1.)

Specify the highest (or lowest) true rating of interest as a percentage. (Example: P ← 100)

PROGRAM INSIDE

Specify the upper criterion level (P1) and the lower criterion level (P2) as percentages.

The program is set for a significance level of 0.1, printed out by step 3 and set by the constant 1.645 in steps 11 and 25. This is easily generalized in the manner of ABOVE and BELOW to take significance level as an input.

The results are printed for a selection of result probabilities. These are based on approximations valid over the probability range from .7 to .995. This range is incorporated in the assumption that all probabilities are taken to be $\leq$ .995 ($t \leq 2.576$), and in the coefficients relating probabilities to t-values (lines 16 and 28). If the probability limits are expanded, the coefficients must be recalculated, and, perhaps, the degree of the polynomial would require increasing.

Step 55 specifies the lowest probability for which results are desired. This is easily changed to an input parameter.

PROGRAM CALC

Input parameters to be set are:

N -- unit size.

K1 -- criterion number of successes in a unit of size N.

K20 -- the true-rating number of successes with which the program starts.

DP -- step decrease in number of successes, starting from K20.

J1 -- sample size with which calculations start.

DJ -- 1+DJ is the number of consecutive sample sizes considered in a calculation run.

SW -- SW≠0 selects a detailed printout; SW=0 selects a minimal printout.

LIMIT -- stops the calculation when the sample size would exceed some number, usually the unit size.

SIGMA -- standard deviation of the distribution based on the criterion value.

X1 -- number of standard deviations over the mean beyond which the criterion value distribution terms are negligible (determined by examination of detailed output).

SIGMA1 -- analogous to SIGMA, but for distribution based on true unit rating.

X2 -- analogous to X1, but representing a lower cutoff on calculation of true-rating distribution terms.

ROW1 -- the first row for which at least one term needs to be calculated. (Row number less one equals the number of successes. Row terms give cumulative probabilities of obtaining this number for various sample sizes. For the distribution based on criterion values, the upper tail is required, and terms are summed in reverse order.)

ROW2 -- the last row for which at least one term requires calculation.

The program can be used with and without the guidance provided by the normal approximation (output of program NORM). An inspection of tables 10 through 13 in the main text shows that selecting the order in which various unit sizes are examined can avoid unnecessary calculation of sample sizes that do not vary with unit size. Note also that only in small regions of the table are adjacent sample sizes in a column within ten samples of each other. Generally DJ should be set to 9, allowing examination of ten sample sizes in one run; a run will then yield only one answer. The true rating, if any, for which one of the ten sample sizes is appropriate is found by starting with the highest success

number possible (the highest K20, considering sample size requirements already determined), and stopping when an answer is found. Table B-1 shows the determination of a sample size requirement of 68 to establish that a unit of size 90, with a true rating of 90.0 percent, is above a criterion level of 84.5 percent. This determination can be made 90 percent of the time with at least 90-percent statistical confidence.

TABLE B-1

PORTION OF A MINIMAL CALC OUTPUT
SHOWING A SAMPLE SIZE REQUIREMENT OF 68

```
UNIT SIZE: 90
PASSES(CRITERION): 76( 84.4 PERCENT)
PASSES(TRUE RATING): 81( 90.0 PERCENT)

   60    2
   61    2
   62    1
   63    1
   64    1
   65    1
   66    1
   67    1
   68    0
   69    0
```

Initially, X1 and X2 are set to some large value, say 20. In the detailed output, we should find that every column has a few zero terms, showing that neglected rows are indeed negligible. The printout, however, is set to suppress rows outside the area of interest; therefore, many rows may be calculated but not displayed. The output of PAR shows which rows are calculated (those from Row1 to Row2, and X1 and X2 are adjusted to keep the calculated rows to only a few more than the required number. This adjustment need not be made very frequently.

The program is set for a significance level of 0.1 and a success probability of 90 percent, both represented by 10-percent tails. In principle, perhaps, we could simply print out all of the matrixes and have the potential of selecting sample sizes for any combination of confidence level and success probability. In fact, however, this method is made practical only by strategems that minimize nonrelevant calculations. If another combination is required, it should be separately calculated. The confidence level is set by the 0.1 in step 29 of CALC. The 0.1 in the next step sets the 90-percent success probability. Steps 13 and 23 limit the detailed printout to terms just short of the 10-percent tail set by the confidence level; an adjustment would be required for a larger confidence level (such as 0.2), and would be desirable for a smaller level.

PROGRAMS HYPRR AND STIR

HYPRR is called by CALC and, in turn, calls STIR. STIR calculates the natural logarithms of the binomial coefficients, and is complicated somewhat by special cases. CALC uses STIR until a hypergeometric term of at least TOL is obtained. Generally a TOL of about 10^{-8} is reasonable. A recursion relation is used to calculate further terms. Terms below 10^{-6} are not calculated now, but assumed to be zero.

PROGRAM PAR

PAR should be used at the start of a run to determine that all parameters have been set correctly. PAR's output should be examined periodically to ensure that a sufficient, but not excessive, number of rows is being calculated.

PROGRAM NORM

NORM requires setting the criterion value to be exceeded (Example: PC ← 0.845) and two coefficients: BETA1 and BETA2. The coefficients are the cutoff value for the tails of a standardized normal distribution for, respectively, the confidence level and the success probability. For example, a confidence level of 0.1 and a success probability of 90 percent represent 10-percent tails, with cutoffs of 1.282. The printout requires specification of the confidence level and success probability (Examples: CONF ← 10 and PROB ← 90). If desired, it is straightforward to cause these to set the coefficients, as was done in programs described earlier.

PROGRAM P

The final program (table A-12) is used to print the correct sample size. The parameter K is set to 1, 2, 3, or 4 to get the correct underlining. The size matrix R is obtained from the same matrix in NORM by replacing, where necessary, the approximated sample size. (By varying the initial value of K, NORM can be used to calculate any of its four matrixes by itself.)